Speech Rate Perception and Speaker Identification in Human-Directed vs. Device-Directed Arabic Speech

This study investigates listeners' perception of human-directed and device-directed speech, focusing on the critical dimensions of speech rate and speaker identification accuracy. Specifically, we investigate how Arabic listeners perceive and process these two distinct speech types. It contributes to the literature by integrating speech rate and speaker identification within a single analytical framework, addressing gaps in previous studies that examined these dimensions separately, and extends the work of scholars such as Cohn and Zellou and colleagues (e.g., Cohn et al., 2024) on human vs. device speech by providing empirical evidence of how these processes manifest in listener perceptions.

Thirty human- vs. device-directed sentences were meticulously controlled for length, speed, content, and speaker characteristics to manipulate the effects of speech direction. Using a rigorous within-subjects experimental design, 78 native Arabic speakers (aged 19-22) participated in two carefully structured tasks to rate the perceived speed of 30 recordings on a 7-point Likert scale and identify whether speakers were addressing humans or devices. Results revealed significant perceptual differences between the two speech types across both dimensions. Device-directed speech was consistently (and correctly) rated as faster than human-directed speech, suggesting that listeners perceive the tempo of technological interactions accurately. Concurrently, human-directed speech was more accurately identified by listeners, indicating that natural human conversation contains distinctive prosodic and rhythmic markers that listeners can readily recognize. Statistical analysis using the Wilcoxon Signed-Ranks Test confirmed these differences were significant with medium effect sizes for both tasks.

These results demonstrate a fundamental trade-off in speech perception: device-directed speech seems to lack the natural variability and adaptability characteristic of human-directed speech. This dichotomy reflects deeper cognitive processes in how listeners categorize and respond to different communicative contexts. The findings have important implications for voice system design, suggesting that incorporating dynamic elements such as natural intonation patterns, contextual adaptability, and rhythm variations could enhance user experience and engagement in technological interfaces. Such improvements would be particularly valuable in applications requiring sustained user interaction, including educational platforms, healthcare communication systems, and sophisticated virtual assistants.

Future research may explore cross-cultural differences in human- vs. device-directed speech perception, examine real-time conversational contexts, and investigate the influence of emotional tone on listener perceptions and preferences, potentially expanding the ecological validity of these findings across diverse linguistic and cultural contexts.

Keywords: Speech perception, human-directed speech, device-directed speech, speech rate, speaker identification, voice system design